

Fast tag compare and bank select in set associative cache

Patent Number: ☐ US5353424
Publication date: 1994-10-04
Inventor(s): PARTOVI HAMID (US); WHEELER WILLIAM R (US); LEARY MICHAEL (US); CASE MICHAEL A (US); BUTLER STEVEN (US); KHANNA RAJESH (US)
Applicant(s): DIGITAL EQUIPMENT CORP (US)
Requested Patent: ☐ JP5225053
Application Number: US19910794865 19911119
Priority Number (s): US19910794865 19911119
IPC Classification: G06F12/08
EC Classification: G06F12/08B10, G06F12/08B22D
Equivalents:

Abstract

A tag comparator and bank selector for a set-associative cache in a computer system operates in a minimum time so that a cache hit or miss signal is generated early in a memory cycle. The data memory of the cache has two (or more) banks, with a tag store for each bank, and the two banks are accessed separately and in parallel using the index (low order address bits) while the tag translation is in progress. Two bit-by-bit tag compares are performed, one for each tag store, producing two multibit match indications, one bit for each tag bit in each tag store. These two match indications are applied to two separate dynamic NOR gates, and the two outputs applied to a logic circuit to detect a hit and generate a bank-select output. There are four possible outcomes from the compare operation: both banks miss, left bank hits, right bank hits, and both banks hit. The later condition indicates a possible ambiguity, and neither data item should be used, so a miss is signalled. The comparator is in large part self-timed using a flow-through design, as distinguished from being enabled on clock edges. Delay elements in the bank select logic allow the banks to be timed against each other, and current limiters are employed to equalize the timing of miss signals, regardless of the number of match lines switching high (which is data dependent). An address producing 19-of-20 match bits will result in a NOR gate output of about the same timing as an address producing no match bits, even though the former will turn on only one transistor to discharge the precharged output node of the NOR gate, whereas the later will turn on twenty paths for discharge. Although a two-way set associative cache is shown herein as an example embodiment, one of the features of the invention is that higher levels of associativity, e.g., four-way and eight-way, are equally well accommodated.

Data supplied from the esp@cenet database - I2

BEST AVAILABLE COPY

ヒラキ

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平5-225053

(43)公開日 平成5年(1993)9月3日

(51)Int.Cl.⁵

G 0 6 F 12/08

識別記号

庁内整理番号

F I

技術表示箇所

E 7232-5B

3 1 0 Z 7232-5B

審査請求 未請求 請求項の数11(全 17 頁)

(21)出願番号 特願平4-307068

(22)出願日 平成4年(1992)11月17日

(31)優先権主張番号 07/794865

(32)優先日 1991年11月19日

(33)優先権主張国 米国 (U S)

(71)出願人 590002873

ディジタル イクイブメント コーポレイ
ション

アメリカ合衆国 マサチューセッツ州
01754メイナード メイン ストリート
146

(72)発明者 ハミド バルトヴィ

アメリカ合衆国 マサチューセッツ州
01581ウェストボロ クレストヴィュー
ドライブ 17

(74)代理人 弁理士 中村 稔 (外6名)

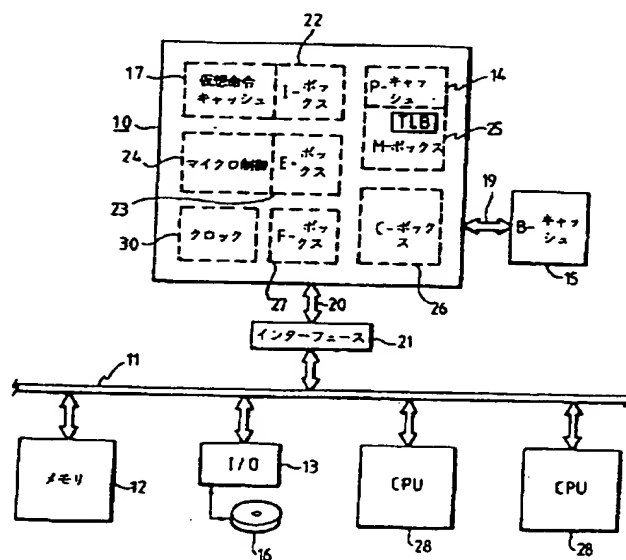
最終頁に続く

(54)【発明の名称】 セットアソシアティブキャッシュにおける高速タグ比較及びバンク選択

(57)【要約】

【目的】 セットアソシアティブキャッシュ操作におけるタグ比較の高速化。

【構成】 プロセッサからタグ値を受け、複数のバンクの中の第1若しくは第2バンクに対する第1若しくは第2記憶タグを受けて複数の第1若しくは第2整合ビット信号をそれぞれ作り出す第1及び第2比較器と、第1若しくは第2整合ビット信号を複数の入力において受け、タグ値と第1若しくは第2記憶タグの全てのビット間で第1若しくは第2整合ビット信号の全てが整合を示したかどうかに応答して第1若しくは第2出力をそれぞれ作り出す第1及び第2NORゲートと、第1及び第2出力を受け、タグ値と第1記憶タグ及び第2記憶タグの整合、非整合に応答して第1若しくは第2バンクに対して第1若しくは第2バンク選択信号を発生するバンクセレクタを備えた複数のバンクを有するセットアソシアティブキャッシュのタグ比較回路。



【特許請求の範囲】

【請求項1】 複数のバンクを有するセットアソシアティブキャッシュのタグ比較を検出するタグ比較回路において、

第1及び第2比較器と、但し、各比較器はプロセッサからタグ値を受け、前記第1比較器は前記複数のバンクの中の第1のバンクに対する第1記憶タグを受けて複数の第1整合ビット信号を作り出し、前記第2比較器は前記複数のバンクの第2のバンクに対する第2記憶タグを受けて複数の第2整合ビット信号を作り出し、タグ値及び記憶タグはそれぞれ複数のアドレスビットを含んでおり、

第1及び第2NORゲートと、但し、前記第1NORゲートは前記第1整合ビット信号を複数の入力において受け、前記タグ値と前記第1記憶タグの全てのビット間で前記第1整合ビット信号の全てが整合を示したかどうかに応答して第1出力を作り出し、前記第2NORゲートは前記第2整合ビット信号を複数の入力において受け、前記タグ値と前記第2記憶タグの全てのビット間で前記第2整合ビット信号の全てが整合を示したかどうかに応答して第2出力を作り出し、

前記タグ比較回路は、更に、前記第1及び第2出力を受けるバンクセクタを備え、このバンクセクタは、前記タグ値が前記第1記憶タグと整合し且つ前記第2記憶タグと整合しないことを前記第1及び第2出力が示した場合には前記第1バンクに対して第1バンク選択信号を発生し、前記バンクセクタは、前記タグ値が前記第2記憶タグと整合し且つ前記第1記憶タグと整合しないことを前記第1及び第2出力が示した場合には前記第2バンクに対して第2バンク選択信号を発生することを特徴とするタグ比較回路。

【請求項2】 請求項1記載のタグ比較回路において、前記バンクセクタは第1及び第2NANDゲート及び第1及び第2インバータを含み、前記第1出力は前記第1NANDゲートの入力に接続され、更に、前記第1出力は前記第1インバータを通じて前記第2NANDゲートの入力に接続されており、前記第2出力は前記第2NANDゲートの入力に接続され、更に、前記第2出力は前記第2インバータを通じて前記第1NANDゲートの入力に接続されており、前記第1及び第2インバータは各々遅延を引き起こして前記NORゲートの前記出力の自己時間調整された経路にマージを与えるタグ比較回路。

【請求項3】 請求項1記載の回路において、前記バンクセクタは、前記第1及び第2バンク選択信号を発生する第1及び第2ゲートを含み、前記第1ゲートは遅延素子を通じて前記第2出力を受け、前記第2ゲートは遅延素子を通じて前記第1出力を受ける回路。

【請求項4】 請求項1記載の回路において、前記NORゲートの前記第1及び第2出力の各々は前記整合ビッ

ト信号が有効となる前にプリチャージされる回路。

【請求項5】 請求項1記載の回路において、前記NORゲートは、前記整合信号を受けるために接続されたゲートと電流制限トランジスタと直列にされたソース・ドレイン経路とを有するNチャンネルトランジスタを備え、前記電流制限トランジスタの各々は比較イネーブル信号によってゲート制御される回路。

【請求項6】 請求項1記載の回路において、前記整合ビット信号は各々、前記第1及び第2比較器の入力において前記記憶タグが有効となる前にプリディスチャージされ、前記第1及び第2出力の一方をディスエイブルし、且つ反対側のバンクをイネイブルすることによってキャッシュフィル動作として前記キャッシュに書き込みを行なう手段を含む回路。

【請求項7】 請求項1記載の回路において、前記比較器の各々は4個の相補トランジスタ対を備えており、各相補トランジスタ対は、出力ノードと電圧源端子の間に直列接続されたソース・ドレイン経路と、前記タグ値若しくは反転されたタグ値、及び、前記記憶タグ若しくは反転された記憶タグに接続された各相補トランジスタ対のゲートとを有し、前記第1及び第2バンク選択信号は第1及び第2ゲートにそれぞれ適用されてプロセッサに対するデータ転送若しくはプロセッサからのデータ転送のために前記第1及び第2バンクを接続する回路。

【請求項8】 複数のバンクを有するセットアソシアティブキャッシュの操作方法であって、

プロセッサから受けたタグ値を前記複数のバンクの第1バンクから受けた第1記憶タグと比較して複数の第1整合ビットを作り出す段階と、

前記タグ値を前記複数のバンクの第2バンクから受けた第2記憶タグと比較して複数の第2整合ビット信号を作り出す段階と、

前記タグ値と前記第1記憶タグの全てのビット間で前記第1整合ビット信号の全てが整合を示したかどうかに応答して第1出力を第1NORゲートにおいて発生する段階と、但しこの発生段階において、前記第1NORゲートは前記第1整合ビット信号を複数の入力において受けるものであり、

前記タグ値と前記第2記憶タグの全てのビット間で前記第2整合ビット信号の全てが整合を示したかどうかに応答して第2出力を第2NORゲートにおいて発生する段階と、但しこの発生段階において、前記第2NORゲートは前記第2整合ビット信号を複数の入力において受けるものであり、

前記第1及び第2出力に応答して第1若しくは第2バンク選択信号のいずれかを発生する段階であって、前記第1バンク選択信号は、前記タグ値が前記第1記憶タグと整合し且つ前記第2記憶タグと整合しないことを前記第1及び第2出力が示した場合に発生され、前記第2バンク選択信号は、前記タグ値が前記第2記憶タグと整合し

且つ前記第1記憶タグと整合しないことを前記第1及び第2出力が示した場合に発生されるような段階と、

前記第1若しくは第2バンク選択信号に応答して、前記第1若しくは第2バンクのいずれかから、前記プロセッサへ若しくは前記プロセッサからデータを結合する段階とを備えることを特徴とする方法。

【請求項9】 請求項8記載の方法において、第1及び第2バンク選択信号を発生する前記段階は、前記第1及び第2出力を第1及び第2ゲートに適用する段階を含み、但し、この適用段階において、前記第1ゲートは遅延素子を通じて前記第2出力を受け、前記第2ゲートは遅延素子を通じて前記第1出力を受けるようにされており、前記発生段階は、更に、前記NORゲートの前記第1及び第2出力を前記整合ビット信号が有効となる前にプリチャージする段階と、前記整合ビット信号の各々を前記記憶タグが前記入力において有効となる前にプリディスチャージする段階を含む方法。

【請求項10】 セットアソシアティブキャッシュを有するコンピュータシステムの操作方法であって、前記システムのプロセッサから受けたタグ値を前記キャッシュの第1バンクから受けた第1記憶タグと比較して複数の第1整合ビット信号を作り出す段階と、前記タグ値を前記キャッシュの第2バンクから受けた第2記憶タグと比較して複数の第2整合ビット信号を作り出す段階と、

前記タグ値と前記第1記憶タグの全てのビット間で前記第1整合ビット信号の全てが整合を示したかどうかに応答して第1ゲートにおいて第1出力を発生する段階と、但し、前記第1ゲートは前記第1整合ビット信号を複数の入力において受ける、

前記タグ値と前記第2記憶タグの全てのビット間で前記第2整合ビット信号の全てが整合を示したかどうかに応答して第2ゲートにおいて第2出力を発生する段階と、但し、前記第2ゲートは前記第2整合ビット信号を複数の入力において受ける、

前記第1及び第2出力に応答して第1若しくは第2バンク選択信号のいずれかを発生する段階であって、前記第1バンク選択信号は、前記タグ値が前記第1記憶タグと整合し且つ前記第2記憶タグと整合しないことを前記第1及び第2出力が示した場合に発生され、前記第2バンク選択信号は、前記タグ値が前記第2記憶タグと整合し且つ前記第1記憶タグと整合しないことを前記第1及び第2出力が示した場合に発生されるような段階と、前記第1若しくは第2バンク選択信号に応答して、前記第1若しくは第2バンクのいずれかから、前記プロセッサへ若しくは前記プロセッサからデータを結合する段階とを備えることを特徴とする方法。

【請求項11】 請求項10記載の方法において、第1及び第2バンク選択信号を発生する前記段階は、前記第1及び第2出力を第1及び第2ゲートに適用する段階を

含み、但し、この適用段階において、前記第1ゲートは遅延素子を通じて前記第2出力を受け、前記第2ゲートは遅延素子を通じて前記第1出力を受けるようにされており、前記第1及び第2ゲートはNチャンネルトランジスタから形成されたNORゲートであり、前記Nチャンネルトランジスタを通じて電流を制限する段階を含む方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、コンピュータ操作、更に言えば、セットアソシアティブキャッシュ操作におけるタグ比較の高速化方法に関する。

【0002】

【従来の技術】CPU性能は、コンピュータハードウェア技術の開発が続く間、メモリスシステムの性能をはるかに上回ってきた。小型の、メインメモリサイズの一部を高速に記憶するキャッシュは、実効メモリアクセス時間を減少させるために使用される。キャッシュは、最も最近用いられたデータの複写物を高速アクセス用として記憶し、また多くのプログラムに関するコンピュータシステム処理速度を大きく改善する。

【0003】あるコンピュータに対するキャッシュ構成を選択するにあたっては、直接マップのキャッシュを用いるか、あるいは連想キャッシュを用いるかということが重要な決定事項となる。(所定アドレスを有する)データ項目(data item)が、キャッシュに出現し得る場所を1つのみ有する場合、そのキャッシュは「直接マップ」である。その項目が、キャッシュ中のどの場所にも位置決めされ得るような場合、そのキャッシュは「完全アソシアティブ」である。その項目が、キャッシュ中の複数位置に対する固定番号(セット)にのみ位置決めされ得るような場合、そのキャッシュは「セットアソシアティブ」である。1つのセットにn個のロケーションが存在する場合、そのキャッシュは「nウェイセットアソシアティブ」である。一般的に、この型のキャッシュで用いられるロケーション数は、1つのセットに2ウェイ若しくは4ウェイ(及び、しばしば8ウェイ)である。連想キャッシュの誤り率は低い、性能上の不都合を引き起こすことがある。

【0004】直接マップキャッシュは最も簡単且つ高速な構成であるが、ある特定のデータ項目が存在し得るキャッシュロケーション数をしばしば限定してしまう。直接マップキャッシュは、所定のインデックスアドレスのデータ項目を記憶し得るロケーションを1つだけしか持っていない。2つ若しくは3つ以上の非常に頻繁に使用されるデータ項目が、直接マップキャッシュ内の同一のキャッシュロケーションにマップされるような場合、これらのデータ項目は、ループ時にプログラムによって周期的に使用され、この結果、キャッシュスラッシングが発生する。各データが使用される場合、使用すべきデータ

が、そのデータの前に存在したデータと置換されるため、メインメモリのアクセスは比較的遅くなってしまふ。キャッシュスラッシングは、しばしば、多くのメインメモリアクセスを引き起こし、プログラムの実行時間を悪化させる。これはまた、理想的な性能を得るためにシステムに必要なシステム相互結合のバンド幅をも増大させる。

【0005】セット連想キャッシュ構成によって、キャッシュにおいて最も最近使用されたデータを発見する公算は増大した。所定のインデックス（低い順位のアドレス）を有するデータ項目を記憶し得るような2つ若しくは3つ以上のロケーションをキャッシュに設けることにより、キャッシュミスは減少し、スラッシングが頻繁に発生することはなくなるであろう。しかしながら、セットアソシアティブキャッシュは、固有的に、直接マップキャッシュよりもその動作が遅い。セットアソシアティブキャッシュによって所定のアドレス値をキャッシュ配列の2つ以上のロケーションにマップすることが可能とされたため、相共通する対応セットアドレスビット（インデックス）を有するような異なるページからの2つ若しくは3以上のアドレスが同時にキャッシュ中に存在し得る。直接マップキャッシュを実行するためのハードウェアは、動作においてはより高速である。しかしながら、セットアソシアティブキャッシュと比較した場合、セットアソシアティブキャッシュにおいては、ヒットであるかミスであるかを判断する間にバス上にデータを駆動できるため、ヒットと判断されなかった場合であってもバス上に正しいデータを駆動することができる。セットアソシアティブキャッシュでは可能性のある整合番号とタグアドレスが比較されなければならない、その後、2つ若しくは3つ以上のバンクからの対応データも選択されなければならないために（タグ比較が完了した後に）、バンク選択という付加的な段階によってその動作は必然的に遅延する。即ち、この付加された時間を許容するためにCPUのサイクル時間がわずかに増大されるため、全システム性能に悪影響を与えてしまう。

【0006】高性能コンピュータシステムは一般に仮想メモリ管理を使用し、仮想アドレスを物理アドレスに変換する間、メモリアドレッシングに遅延を生じさせる。マイクロプロセッサ用のオンチップキャッシュは、キャッシュをアクセスしてメモリ参照データを含むかどうかを確認することができるようになる前に、通常はオンチップ変換バッファによってアドレス変換に必要とされるその応答時間内に抑制される。変換バッファがページフレームアドレスを仮想メモリ管理システム内に作り出し、その後、セットアソシアティブキャッシュの比較及びバンク選択操作に対する性能要求を更に満たすまでは、タグ比較の開始が待機される。

【0007】CPUサイクル時間における改善と調和させるべく、回路に様々な改善を施し、セットアソシアテ

ィブキャッシュの比較操作を高速化している。目標は、1つのCPUサイクルで（階層メモリ内の）第1のキャッシュをアクセス可能することであり、第1のキャッシュにおいてキャッシュミスが発生した場合には、1つ若しくは2つのサイクル内に第2のキャッシュに対するアクセスを完了することである。これには、アクセスサイクルのできるだけ早い時期に、つまり、タグが利用可能となった後にできるだけ早く、ミスが検出されることが必要である。半導体製造及び回路設計技術においては、100Mhz（及びそれ以上）のクロックが利用可能であり、10n秒CPUサイクル時間を提供する。

【0008】Suzuki等による論文、「A 19ns Memory」、ISSCC 87, p.134 には、高速SRAMキャッシュ用のタグメモリが開示されている。このタグメモリは、CMOS整合回路と、ヒット信号を発生するNMO S NORゲートを備えた比較器を用いている。しかしながら、Suzuki等による比較回路は、バンク選択信号を高速に与えるものではなく、また、そのヒット信号は、データに依存したもの、つまり、そのタイミングはタグアドレスで整合したビット番号とは異なるようにして依存する。更に、Suzuki等による回路は、もしこれがセットアソシアティブキャッシュに使用された場合には、ヒット信号が共に主張されるためにクロックエッジ(clock edge)に対して状態を調整しなければならない。また、Suzuki等によるキャッシュは直接マップされるため、複数のヒットを検出することはない。複数のヒット検出を有するキャッシュは、Ooi等による論文、「Fail-Soft Circuit Design in a Cache Memory Control LSI」、ISSCC87, p.103 に開示されているが、この回路はこの機能を提供するために、更に別の論理装置を必要とするため、比較操作は更に遅延してしまう。

【0009】

【発明の概要】本発明の一実施例によれば、コンピュータシステムのセットアソシアティブキャッシュに対するタグ比較が最小時間で動作するよう構成されるため、キャッシュミス信号はメモリサイクルの初期に発生される。キャッシュのデータメモリは2つ（若しくは3つ以上）のバンクを有し、各バンクに対してタグ記憶を備える。これら2つのバンクはタグ変換がTLBで進行する間、インデックス（低順位アドレスビット）を用いて個別且つ並列にアクセスされる。ビット対ビットの比較が2度行われる。一方はタグ記憶の各々に関するものであり、CPUからの物理アドレスのタグフィールドが記憶値と比較され、2つのマルチビット整合表示を作り出し、1つのビットが各タグ記憶の各タグに対応する。これら2つの整合表示は2つの個別のダイナミックNORゲートに適用され、これら2つ出力はバンク選択出力の発生のために論理回路に適用される。比較動作には4つの可能な結果が存在する。即ち、両方のバンクがミスを

発生した場合、左バンクがヒットを発生した場合、右バンクがヒットを発生した場合、両方がヒットを発生した場合である。両方のバンクでヒットが発生した状態は、どのような可能性も可能であることを示しており、いづれのデータ項目も使用されず、従って、ミスが信号合図されるべきである。本発明の比較器の特徴は、両バンクにおけるヒットの状態が検出され、付加的な論理回路や遅延を生じることなく、ミスが信号合図されるということである。他の特徴は、クロックエッジでイネーブルされるものと異なり、フロースルー設計（それらのバンクは互いに時間調整されている）を用いて、ほとんどの部分を自己時間調整していることである。この方法では、最悪のタイミングをも許容するよう設計されたクロックを待たずに各段階において比較及びバンク選択動作が完了されるため、全平均速度が強化される。バンク選択ロジックにおける遅延素子によってバンク間で互いにタイミングを取ることが可能とされ、高状態に切り代わる整合ラインの数（ミス比較の数、1〜20、に依存する）に拘わりなく、ミス信号のタイミングを均一化するために電流制限が使用される。故に、19オブ20整合ビットを作るアドレスは、0オブ20整合ビットを作り出すアドレスとほぼ同じタイミングのNORゲート出力を生じさせ、これは、前者がトランジスタの一方だけをオン状態としてNORゲートのチャージされた出力ノードをディスチャージしたとしても、後者はディスチャージのために20個の経路をオン状態とするだろう。本明細書では一例として2ウェイセットアソシアティブキャッシュを示したが、本発明の特徴の中の1つはより高レベルのアソシアティブ、例えば4ウェイ及び8ウェイにも同様に良好に適用できるであろう。

【0010】

【発明の詳細な説明】図1を参照すれば明かなように、本発明の一実施例によれば、本発明のキャッシュ制御特性を用いるコンピュータシステムは、CPUチップ若しくは、William Wheeler等によって1990年6月29日に出願され、本発明の譲受人であるDigital Equipment Corporationに譲渡された継続中の出願、シリアル番号547,824号に開示された型のモジュール10を有する。この実施例において、CPU10は、システムバス11によって、システムメモリ12及びI/O素子13に接続されている。より好ましい実施例において、本発明によるキャッシュ制御は、シングル集積回路上に形成されたCPU10に対するオンチップキャッシュとして使用されるものであるが、以下に述べる幾つかの概念は、シングル回路ボード若しくは多重ボード上に載置されたチップセットとして実行することができる。命令若しくはデータにフェッチを行った場合、CPU10は、内部、即ち第1キャッシュ14にアクセスする。このキャッシュ14は、以下に述べる本発明のキャッシュ制御概念を用いる。第1のキャッシュにおいてメモリ参照が発見されな

かった場合、より大きな外部（オフチップ）若しくはバックアップキャッシュ15にアクセスする。これは階層メモリを提供するものであり、まず、第1のキャッシュ14が、次に、バックアップキャッシュ15が、次に、メインシステムメモリ12が、更に、一般には、ディスクメモリ16がその後続く。このディスクは、オペレーティングシステム（つまり、ソフトウェア）を用いてページング及び仮想メモリ管理のためにアクセスされる。仮想メモリ構造はディスク16上に記憶され、オペレーティングシステムによって物理メモリ12に運ばれたページを用いて、メモリ12で最も使用される可能性が高いページを保持する。CPU10の付加的なキャッシュ17は、物理アドレスの代わりに仮想アドレスを使用して命令のみを記憶する。物理アドレスは、第1及びバックアップキャッシュ14、15をアクセスするために用いられ、バス11上及びメモリ12で使用される。CPU10が命令にフェッチした場合、まず、仮想命令キャッシュ17が調査され、キャッシュミスが発生している場合には、そのアドレスは物理アドレスに変換されて第1のキャッシュ14が調査される。命令が第1のキャッシュに存在しない場合、次にバックアップキャッシュ15が調査され、このバックアップキャッシュでキャッシュミスが発生した場合に、メモリ12に対してアクセスが行われる。データも同様に階層メモリシステムでアクセスされるが、本実施例においては、データに対しては仮想命令キャッシュに対応するものは設けられていない。第1キャッシュ14はその大きさはより小さいが、バックアップキャッシュ15より高速であり、また、第1キャッシュ14の内容はバックアップキャッシュ15の内容のサブセットである。仮想命令キャッシュ17は、命令がメモリからフェッチされた場合のキャッシュフィル操作(cache fill operation)を除けば、CPUからキャッシュ17への書き込みは存在しないという点で、また、キャッシュ17の内容はキャッシュ14、若しくは15の内容のサブセットである必要はないという点で、その内容がキャッシュ14、若しくは15の内容のサブセットであるかもしれない他の2つのキャッシュ14、15の動作とは異なる。

【0011】CPU10がバス19を通じてバックアップキャッシュ15にアクセスし、システムバス11に対するアクセスのためのCPUバス20からは分離されているため、バックアップキャッシュ15に対するキャッシュコントローラはCPUチップ内に含まれる。CPUバス20及びシステムバス11は64ビットの両方向多重アドレス/データバスであり、リクエスト、グラント(grant)、コマンドライン等を含む制御バスが加えられている。しかしながら、バス19は64ビットデータバスと、別個のアドレスバスとを有する。システムバス11は、インタフェースユニット21によって、CPUバス20に相互接続される。インタフェースユニット21

は、CPU10、及びCPUバス20上の他の素子によるアービトレイトアクセスに機能する。

【0012】CPU10は命令ユニット22（I-ボックスと呼ばれる）を含む。命令ユニット22は、マクロ命令（機械レベル命令）をフェッチし、1サイクルにつき1つのマクロ命令をデコードし、オペランド指定子

（operand specifiers）を分析（parse）し、その後、そのオペランドフェッチを開始する。命令によって指令されたデータ若しくはアドレス操作は、レジスタファイル及びALUを備える実行ユニット、即ち、E-ボックス23によってなされる。CPUはマイクロコードによって制御され、マイクロプロセッサ及び制御記憶装置を備えるマイクロ命令制御ユニット24を用いて、マクロ命令の実行に必要なマイクロ命令のシーケンスが発生される。メモリ管理装置、即ちM-ボックス25は、命令読出、及びデータ読出用のリクエストを命令ユニット22からデータ読出若しくは書き込みリクエストを実行ユニット23から受け取る。アドレス変換はTLBを用いてM-ボックス25で行われ、仮想アドレスから物理アドレスを発生させる。M-ボックスはP-キャッシュに対してメモリリクエストを発生し、P-キャッシュミスの場合には、キャッシュコントローラ、即ちC-ボックス26を介して、バックアップキャッシュ15にリクエストを送る。このキャッシュコントローラ26は（以下に述べる）P-キャッシュに対するキャッシュコントローラとは別であり、P-キャッシュミスの場合はバックアップ（第2レベル）キャッシュ15に対するアクセスを処理し、また、バックアップキャッシュミスに対してはバス20を介してメインメモリ12にアクセスを行なう。オンチップ浮動少数点プロセッサ（F-ボックスと呼ぶ）は、浮動少数点と整数の掛け算命令のための実行ユニットであり、実行ユニット23からオペランド及びコマンドの受取り、且つ実行ユニットへその結果を戻す。

【0013】本発明のキャッシュ制御特性は、様々な型のCPUで使用することが可能であるが、本明細書に開示したCPUの一例はVAX™アーキテクチャであり、これは、LevyとEckhouseによる「Computer Programming and Architecture: The VAX」、第2版、Digital Press、1989に記述され、本明細書に参考として組み入れられている。CPU10はLevy等によるテキストで、若しくは、Sander、Uhler及びBrownに付与され、本発明の譲渡人であるDigital Equipment Corporationに譲渡された米国特許第5,006,980号によって開示されているような、VAX™タイプであろう。

【0014】本明細書に記述したCPUの一実施例はVAXアーキテクチャであるが、それにも係わらず、本発明のキャッシュ制御の特性は80386若しくは68030タイプ等の他のアーキテクチャに対して構成されたプロセッサにおいても同様に有用である。また、本明細

書で開示されたCISCタイプのコンピュータではなく、MIPS、R3000や、1990年6月29日に出願された継続中の出願、シリアル番号547,630号、に開示された改良された64ビットRISCアーキテクチャのようなRISCタイプのコンピュータにも有用である。

【0015】付加的なCPU28は、マルチプロセッサシステムのシステムバス11にアクセスすることができる。付加的なCPUの各々は、それらがCPU10と同じ設計ならば、自身のCPUチップ10、第1キャッシュ14、バックアップキャッシュ15、及びインタフェースユニット21を含むことが可能である。また、これらの他のCPU28は異なる構成であってもよいが、互換性のあるバスプロトコルを実行してメインシステムバス11にアクセスする。これらの他のCPU28はメモリ12にアクセスすることができ、従って、キャッシュ14若しくは15のデータブロックを排除することができる。CPU28がメモリ12のあるロケーションに書き込みを行い、第1キャッシュ14（若しくは、バックアップキャッシュ15）に重複が発生した場合、キャッシュ内のこのロケーションにおけるデータはもはや有効ではない。故に、CPU10以外のソース（他のCPU28）からメモリ12への書き込みが存在した場合、キャッシュ14及び15のデータブロックは無効とされる。第1のキャッシュ14は「ライトスルー」原理で動作するが、それにも係わらず、キャッシュ15は「ライトバック」原理で動作する。第1キャッシュ14に生じたロケーションに対してCPU10による書き込みが実行された場合、キャッシュ14、更に、バックアップキャッシュ15に（状態に依存して、しばしばメモリ12にも）データが書き込まれるが、この動作タイプは「ライトスルー」である。CPU10がバックアップキャッシュ15のロケーションに対して書き込みを実行した場合、その書き込みをメモリ12に向ける必要はないが、その代わり、システムの他の素子（CPU28のような）がそのデータを必要とした場合（つまり、メモリのこのロケーションにアクセスを試みた場合）若しくは、キャッシュのキャッシュラインがキャッシュ15から置換され（配置が解除され）た場合にのみ、メモリ12にライトバックされる。

【0016】図2を参照すれば明かなように、図1のCPUチップ10のクロック発生器30は重複しない4つのクロック、phi1、phi2、phi3及び、phi4発生し、これらのクロックはマシンサイクルの4つの位相を決定するため使用される。ある一つの実施例において、マシンサイクルは公称で10n秒（100MHz）であるため、クロックphi1等はそれぞれ2.5n秒で動作する。第1キャッシュ14は、単一のマシンサイクル内でM-ボックス25に対して読出データを戻し、キャッシュミス若しくはキャッシュヒットは1つの

メインサイクルよりも非常に小さいマシンサイクル内で決定される。この結果、ミスの場合にヒット若しくはコマンドがバックアップキャッシュに進むことができる場合には、データアクセスは完了され得る。外部バス19及び20はCPUのマシンサイクルよりも3つの時間だけ長いバスサイクル上で動作するため、この例では、この外部バスサイクルは図2にも示されているように公称で30n秒である。バスサイクルは、バスサイクルの4つの位相を定めるためにクロック発生器30によって作り出された4つの重なり合うクロック、Phi1、Phi2、Phi3及び、Phi4によって同様に定義される。しかしながら、システムバス11は、バス20のバスサイクルの約2倍の長さを有するより長いバスサイクル、例えばおよそ64n秒で動作し、このバスサイクルはCPU10及びバス20とは非同期である。システムバス11のタイミングサイクルはインタフェースユニット21のクロック発生器によって制御される。

【0017】CPUチップ10の内部構成は、図1にその概観が示されている。命令ユニット22は仮想命令キャッシュ17を含む。この仮想命令キャッシュ17は2Kバイトサイズの専用命令ストリームオンリキャッシュ(dedicated instruction-stream-only cache)であり、この命令キャッシュは、ある実施例においては、最も最近使用された命令ストリームのブロックを記憶し、物理アドレスではなく仮想アドレスを使用する。物理アドレスはキャッシュ14及び15(及び、メインメモリ12)にアクセスするために用いられる。即ち、仮想命令キャッシュ17にアクセスするためのアドレスはメモリ管理ユニット25において他のメモリ参照に関してなされるように、TLB31でアドレス変換を行なう必要はない。これらの命令は命令キャッシュ17から、そこからの命令がデコードのために採用されるプリフェッチキュー(prefetch queue)にロード(load)される。

【0018】マイクロ命令制御ユニット24は制御記憶装置からフェッチされるべきマイクロワードを選択するマイクロシーケンサを含む。制御記憶装置はROMであり、マイクロシーケンサによって発生されたアドレスに応答してマシンサイクル毎にマイクロコードワードを作り出す。マイクロシーケンサは命令ユニット22からエントリポイントアドレスを受け取り、マクロ命令によって命令されたマイクロルーチンを開始する。(マイクロ命令制御ユニット24の制御下にある) 実行ユニット23はマクロ命令の実際の「作業」を実行する。実行ユニット23に含まれるレジスタファイルは汎用目的のレジスタ、プログラムカウンタ、メモリデータレジスタ、一時(temporary)若しくは作業レジスタ、及び状態レジスタを含む。実行ユニット23はまたALU及びシフタ(shifter)を含み、マイクロシーケンサから受け取ったマクロ命令によって指定された操作を実行する。

【0019】メモリ管理ユニット25は命令ユニット2

2から(命令流れ、及びデータ流れの両方)及び、実行ユニット23から(データ流れのみ)読出リクエストを受け、メモリ管理ユニット25は命令ユニット22若しくは実行ユニット23の両方にメモリ読出データを送信する。メモリ管理ユニット25はまた書き込み/記憶リクエストを実行ユニット23から受け、同様に無効とし、第1キャッシュ14を満たし、キャッシュコントローラユニット26からのデータを返す。メモリ管理ユニット25はこれらのリクエストを裁定し、現在処理できないリクエストを待ち行列に入れる。一旦、リクエストが開始されると、メモリ管理ユニット25は変換バッファ若しくはアドレスキャッシュ31を用いてアドレス変換を行い、仮想アドレスを物理アドレスにマッピングする。アドレスキャッシュ31におけるこの検索には、

(もしヒットが存在するならば) 1マシンサイクルも必要としない。TLB31でミスした場合、メモリ管理回路によってメモリのページテーブルからページテーブルエントリが読み出され、ミスしたアドレスを挿入するためにTLBフィルが実行される。このメモリ管理回路はまた全てのアクセスを調査し、ページ保護機能等を実行する。以下で述べるように、メモリ管理ユニット25によって参照されたP-キャッシュ15は2ウェイセットアソシアティブライトスルーキャッシュであり、1つのブロックと、32バイトのフィルサイズ(キャッシュライン)を備える。P-キャッシュ状態はバックアップキャッシュ15のサブセットとして保持される。メモリ管理ユニット25によって受け取られたが第1キャッシュ14でミスが発生したようなメモリリクエストは、キャッシュコントローラユニット26に送られて実行される。メモリ管理ユニット25は無効を受けて、キャッシュコントローラユニット26からのデータを満たす。

【0020】キャッシュコントローラユニット26はバックアップキャッシュ15に対するコントローラであると共に、外部CPUバス20に対するインタフェースである。キャッシュコントローラ26は読出リクエストを受け、メモリ管理ユニット25から書き込みを行い、第1キャッシュ14のフィル及び無効をメモリ管理ユニット25に送る。キャッシュコントローラユニット26は、第1キャッシュ14が無効によってバックアップキャッシュからのサブセットとして保持されることを確実なものとする。キャッシュコントローラユニット26はバス20からキャッシュコヒーレンシーランザクションを受け、無効及びライトバックによってこれに対して適当に応答する。図1のシステムのキャッシュコヒーレンシーはオーナーシップの概念に基づいたものである。即ち、六ワード(hexaward)(16ワード若しくは32バイト)ブロックのメモリを、メモリ12、若しくはバス11上のCPUのバックアップキャッシュ15の両方が所有することができ、マルチプロセッサシステムにあっては、これらのキャッシュの中の1つだけ若しくはメ

メモリ12が所定時間にその六ワードブロックを所有することができ、このオーナーシップはメモリ12及びバックアップキャッシュ15両方の各六ワードに対するオーナーシップビットによって表示(1は所有、0は非所有)される。バックアップキャッシュ15のタグ及びデータは共にオフチップRAMに記憶されており、システム要求の必要に応じて選択されたサイズ及びアクセス時間を有する。バックアップキャッシュ15は例えば128K~2Mバイトのサイズである。28n秒のアクセス時間を有する場合に、CPU10に対して10n秒マシンサイクルと仮定すれば、3マシンサイクルでキャッシュを参照することができる。

【0021】図3を参照すれば明かなように、ライン33及び34上の物理アドレスはTLB312からの出力(仮想アドレス35からのバイパスライン34を含む)において、ビット<31:9>を含む22ビットのページフレーム番号36を備え、各ページが512バイトであるような4Mページの中の1つを選択する機能を持つ。6ビットページオフセット37は、1つのページ中の1オブ64(1-of-64)四ワード(quadwords)を選択するビット<8:3>を含む。3ビットのバイトアドレス38、<2:0>は、1つの四ワード中の1オブ8バイトを選択する。物理アドレスに対するこの区分けは、512バイトのページに対するものである。1K若しくは4Kバイトのような、他のページサイズで実施することも可能であるが、この場合、ページフレーム番号はより小さなビットとなり、ページインデックス37はそれに対応してより大きなものとなる。PFNフィールド36、ビット<31:9>は、変換のためにMボックス25のTLB31を通過した仮想アドレスの一部であるが、これらのビットは遅延され、ライン34を経由してTLB31を回避したビット<8:0>と比較され、P-キャッシュ14をアクセスするためにすぐに利用可能となる。実施例の第1キャッシュ14の構成では、図3に示したように、タグ40は20ビット幅(ビット<31:12>)であり、インデックスフィールド41は7ビット幅(ビット<11:5>)である。インデックス41はこのように、アドレスサイクルの早期に利用可能なく8:5>を部分的に形成するものである。これは、これらのビットがTLB31を回避するためであり、ビットは部分的に、<11:9>、サイクルのより遅い時期に利用可能であるため、ライン33上のTLBからPFN36のより遅いビットとして到達する。ビット<4:3>からなる4ワード選択フィールドは、ライン34上のページオフセットの初期部分に存在し、データ記憶配列の4つの4ワードの中から選択するために使用される。図4を参照すれば明かなように、第1キャッシュ14は左バンク43Lと右バンク43Rといった2つの識別バンクに構成されて(論理形態で)示されている。各バンクのそれぞれにタグ記憶44若しくは44R及

び、データ記憶45L若しくは45Rが存在する。各バンクに128個のインデックス、即ち行46が存在し、各インデックスは8個の4ワード47のデータ(各バンクに4個)を含み、64ビット若しくは8バイトの4ワードを備える。この結果、この実施例において、第1キャッシュ14に対するキャッシュサイズは128×8×8、即ち8Kバイトのデータ(更に、タグ等を加えて)である。行(インデックス)デコードは128の中から1つ選択するのであるが、これは2段階で実行される。第1に、キャッシュの各バンクの8つの部分の各々において、インデックス31の初期のビット<8:5>によって駆動された8つの1オブ16(1-of-16)デコーダ48のセットを用いて16個の中から1個が選択される。次に、インデックス41の後方のビット<11:9>が到達し、これら後方のビットに回答して1オブ8セクタ49のセットによって第2の選択がなされる。初期ビット<4:3>を用いてデコーダ50のセットによって4個の中から1個を選択し、インデックスに対する各バンクに対して4ワード47の1つを選択する。読出にあっては、初期ビット<8:5>が使用されているため、読出データの8つのセットをデコーダ48の出力において初期に(各バンクにおいて)利用することができる。この結果、TLBアクセスが達成される間に少なくとも部分的にキャッシュアクセスが発生するため、全データアクセスが高速化される。インデックス46が選択された場合、タグ記憶44L若しくは44Rがデータ記憶45L若しくは45Rとともに同時にアクセスされる。128個のインデックス、行46の各々は、各バンクに20ビットのタグ51を含む。データ記憶がアクセスされると同時にタグ51を選択するために、同一のデコーダ48及び49が使用される。後方ビット処理に必要なのはたった3ビット、8個の中の1個、のデコードであるため、また、初期のビットはデコーダ48の8つの出力を通じて配列アクセスを行なうために既に使用されているため、バス33上でPFN36をTLB31が利用できるようになったすぐ後に、各バンクに対するタグ記憶の出力52L若しくは52Rにおいて20ビットタグ51を利用することができる。

【0022】図5に、図4のデータ及びデータ記憶を形成するメモリセル配列及びセクタの構造が、セクタ48、49及び50の重なりを示すような図で示されている。Hamid Partovi 及びMichel Case によって1990年4月11日に出願され、本発明の譲受人であるDigital Equipment Corporation に譲渡された継続中の出願、第508,082号に開示されているように、配列構成はスーパービットライン(super bit lines)を用いる型をしている。図5から明かなように、1オブ16セクタ48の各々はセル配列の16個のメモリセル53に接続されており、1オブ4セクタ50の1つはこれらの1オブ16セクタ48の中の4個に及ぶ。1オブ8セクタ49

は1オブ4セクタ50の中の8個に及び、データ（更に、パリティ、有効ビット等）に対しては、左及び右バンク45L及び45Rの各々にこれらの1オブ8セクタ49が64個存在する。1オブ8セクタ49だけが後方ビット<11:9>を使用するため、配列アクセスは1オブ16及び1オブ4セクタ48、50を通じて初期ビットを用いて達成される。タグ記憶44L及び44Rは一般にはデータ記憶と同様であるが、1オブ4セクタ50は必要とせず、各タグに対して20ビット若しくは20個の1オブ8セクタ49のセットだけを含み、バンクに1つに対する1つの20ビットタグが存在し且つ2つバンクが存在する。

【0023】図6を参照すれば明かなように、タグ記憶の単一インデックス、即ち、単一行（キャッシュライン）とデータ記憶は、共に、627ビット値54であり、左及び右バンクの各々に、1ビットのタグパリティ55を有する20ビットタグ51と、各バイトに1つのパリティビットを備えるキャッシュデータの4つの四ワード72（ $4 \times (64 + 8)$ ）と、更に、4つの有効ビット56（データ47の各四ワードに対して1つ）を含む。配置ポインタビット57は各インデックスに対して含まれる。タグパリティビット55はタグ51と共にアクセスされ、タグのパリティはタグデータバスと同時に動作する回路によって調査される。しかしながらこれは従来の構成であるため、本明細書に開示及び記述していない。同様に有効ビット56は、データアクセスに関して図示された汎用形態と同一の回路によって、データアクセスと同時に論理回路によって設定若しくは調査される。配置ポインタビット57は、最後に使用されたものでない（not-last-used）置換アルゴリズムを実行するために用いられる。キャッシュミスが発生した場合はその結果が表示され、このインデックス46を共有する2つの可能な4つの四ワード47のセットの中の1つがキャッシュに重ね書きされるであろう。アルゴリズムは左及び右という2つのバンクに対するビット57の内容を用いて実施され、この決定がなされるが、これはキャッシュリフィル（refill）動作の間に発生するものであり、キャッシュヒット若しくはミス決定と関連する本発明とは関係ない。

【0024】図4において、タグ記憶44L及び44Rからのライン52L及び52R上のタグ出力は整合（比較）回路58L及び58Rに付与され、また、これらの回路はTLB31からライン33を通じてタグ40を受ける。これらの回路58L及び58Rは、NOR論理回路59L若しくは59Rに対する入力として20個の整合ビットのセットを作り出し、それはまた、バンク選択回路61に対する入力としてライン60L及び60R上にミス信号を作り出す。バンク選択回路は、ライン62I上に左バンク選択信号、あるいは、ライン62R上に右バンク選択信号を作り出すが、両方に作り出すという

わけではない、またこれらの信号は、ライン64L若しくは64Rをメモリ管理ユニット25に向かうデータバス65に結合するためにゲート63L及び63Rによって用いられる。

【0025】図7にメモリサイクルのタイミングが示されており、本発明の一実施例におけるキャッシュ比較及びアクセスが表されている。第1のサイクルのphi1、phi2及び、phi3の一部を占有する間、時間66aにおいてライン33上で利用可能なタグを用いてTLB31の検索が発生する。この時間間隔66の初期部分66bの間、セクタ48において初期ビット<8:5>を用いてデコード動作が発生し、第1のサイクルのphi1及びphi2を占有する。時間66aにおいて、ライン33で後方ビット及びタグを利用することが可能となり、時間間隔67の間に配列のワードラインが駆動される。この時間67の中の間隔67a（通常はphi4で）において、1オブ8セクタ50においてデコードが完了し、回路58L及び58R、59L及び59R及び61においてバンク選択動作が完了する。キャッシュヒットが読出動作で検出されたと仮定した場合、間隔67aの比較動作の最後に、ライン64L若しくは64R及びゲート63L若しくは63Rを介してキャッシュデータ記憶45L、若しくは45Rの所定の左、若しくは右バンクの出力を用いて、バス65が駆動される。読出しではなく書き込み動作に対して信号合図がなされると、ワードラインが駆動される間隔67の中の第2のサイクルのphi1の間隔68の間に、所定の左若しくは右バンクに対する書き込みが達成される。新たなアクセスサイクルを他のTLB検索間隔69の間に開始することができる。このアクセスに対する間隔69bは初期ビットに対するデコードであり、丁度完了したアクセスサイクルの間隔66及び66bに対応する。

【0026】図8に、本発明の一実施例による各回路58L、58R、NOR回路59L、59R、及びバンク選択回路61が詳細に示されている。左バンクに対しては20個のタグビット比較回路70のセットが用いられ、右バンクに対しては20個のタグビット比較回路71のセットが同様に用いられる。ライン33を経由してTLB31から受け取られたタグ40は、ライン33₀～33₁₉上のこれらの比較回路70及び71の全てに対する1つの入力である。各比較回路70はライン52Lの20個のセットの中の1つに存在する左バンクタグデータ記憶配列スライス（left bank tag data store array slices）の1つから、デュアルレール出力の1つを受け取る。同様に各比較回路71は、ライン52Rの20個のセットの1つに存在する右バンクタグデータ記憶配列スライスの1つからデュアルレール出力の1つを受け取る。各比較回路70若しくは71は、直列接続されたPチャンネルトランジスタ72、73及びNチャンネルトランジスタ74、75を1つのグループに、更に、直

列接続されたPチャンネルトランジスタ76、77及びNチャンネルトランジスタ78、79を第2のグループに有するようなCMOS排他OR回路である。整合回路70、71からの出力はNORゲート59L及び59Rに向かうライン80上に存在する。ライン52L₀が高、ライン52L₀バーが低、及びライン33₀が高の場合、トランジスタ74及び75はオン状態であり、トランジスタの他の全ての対は1つのオフ状態のトランジスタを有するため、この特別な整合回路70に対する出力ライン80は低状態に引き寄せられ整合を表示する。同様に、ライン52L₀が低、ライン52L₀が高、及びライン33₀が低の場合、トランジスタ78及び79はオン状態であり、トランジスタの他の対は全て1つのオフ状態のトランジスタを有するため、この回路70に対する出力ライン80もまた低状態に引き寄せられ、整合を表示する。他の2つの可能な組合せ(ライン52L₀が低、ライン52L₀バーが高、若しくは、ライン52L₀が高、ライン52L₀バーが低)によってこのライン80はトランジスタ対76、77、若しくはトランジスタ対72、73を通じて高状態に引き寄せられ、このタグビットに対して非整合を表示する。ライン52L及び52Lバー、52R及び52Rバーは全てphi2の間にV_{dd}にプリチャージ(precharge)され、phi3で駆動されるが、ライン80は全てphi2の間にグラウンドに対して予めプリディスチャージ(predischage)され、phi4で駆動される。ライン52L、52Lバー等の全てがV_{dd}にプリチャージされている場合、ライン80は全てプリディスチャージされる(グラウンドに落とされる)。なぜなら、グラウンドに対する経路はこのビットに対するライン33が1であるか0であるかに係わらずトランジスタ72、73若しくは76、77によって確立されるためであり、ライン80のこのプリディスチャージは回路の自己タイミング操作の一部である。

【0027】各バンクに対する20個のライン80、即ち、図8の整合回路58L及び58Rの出力は、左及び右バンクに対するNチャンネル59L及び59Rを備えた20個のトランジスタ81若しくは82の2つのセットであるようなゲートに接続される。これらのゲートはphi2、即ち、プリチャージ間隔の間に、低状態に保持され、非整合が発生した場合は1つ若しくは2つ以上がphi4で高状態となる。トランジスタ80若しくは81の各セットのドレインは出力ライン60L及び60Rに接続される。これらの出力ラインは、Pチャンネルトランジスタ85によって、phi3の間にV_{dd}に別個にプリチャージされる。トランジスタ81及び82のセットのソースは電流制限用トランジスタ86、87によってV_{ss}に接続される。トランジスタ86、87のゲートはライン88の比較イネーブル信号によって駆動され、phi4でしばしば高状態となる。この左バンクに対するライン80の1つ若しくは2つ以上がphi4

の間に高状態である場合、phi4の間、出力ライン60Lは低状態に引き寄せられ、ミスを表示する(ライン33上のタグ40は、ライン52L上の記憶タグ51と等しくない)。同様に、この右バンクに対するライン80の1つ若しくは2つ以上がphi4の間に高状態である場合、phi4の間、出力ライン60Lは低状態に引き寄せられ、ミスを表示する(ライン33上のタグ40はライン52R上の記憶タグ51と等しくない)。ライン60L若しくは60Rは、全てのトランジスタ81若しくは82がオフ状態(バンクに対する20個のライン80の全てが低状態)である場合、phi4で高状態であり、左若しくは右バンクに関してヒット若しくはタグ整合を表示する。ライン60L若しくは60Rはインバータによって駆動されたゲートを有する高インピーダンスPチャンネルトランジスタ89によって高状態に保持され、また、これらのトランジスタによって低状態に保持することができる。

【0028】バンク選択回路61は図8に示されているように、2個のインバータと共に2個のNANDゲート90、91を含む。ゲート90はライン60L上の反転ミス(アクティブラウ(active low))出力、及びライン60R上のミス(アクティブラウ)出力を受け、ライン62R上に右バンク選択出力(アクティブラウ)を作り出す。ゲート91はライン60L上のミス(アクティブラウ)出力、及びライン60R上の反転ミス(アクティブラウ)出力を受け、ライン62L上に左バンク選択出力(アクティブラウ)を作り出す。インバータ92、93は各NANDゲート90、91の入力の1つに対して遅延を与える。ライン60L若しくは60Rが、ミス信号の遅延ウィンドウ(インバータ92若しくは93によって形成される)以内に反対側のバンクによって作り出された他のライン60L若しくは60R上にミスを信号合図しなかった(つまり、高状態である)場合、そのバンクはヒットする。インバータ92、93はバンク選択論理の実施し(つまり、所定のアドレスに対して、どのバンクをアクセスすべきかを選択する)、また、ライン60L及び60R上の信号がエラーを起こした場合(やがては、スキュー(skew)する)にマージン(margin)を与える遅延素子としても機能する。例えば、このマージンはバンクの一時的な(且つ、エラーを伴って)選択を引き起こすが、バンクは両方とも最後にはミスしてしまうようなエラーを補うことが可能である。図9の(a)及び図9の(b)に示されているように、インバータに遅延が存在せず、ライン60L及び60R上にミス信号が現れている状態を考えてみる。ミス信号は、左バンクに関しては時間 t_1 でアクティブラウにドロップするが、ライン60Rに関しては t_2 までドロップしない。ゲート91はいづれの時点においても両入力において(低出力を作り出す必要があるような)高状態とならないため、図9の(c)に示されているように左バンク選択ラ

イン62Lはずっと高状態のままである。右バンク選択ラインは、図9の(d)に示されているようにエラーを伴うような実行を禁止するが、ライン60Rが高状態、ライン60Lが低状態であっても、 t_1 及び t_2 の間に誤った「ヒット」を示す。これは、この時間間隔の間、ゲート90が低状態の出力を作り出すような2個の高状態入力となるからである。図9の(d)のこの誤った「ヒット」信号、ゲート63Rを介してバス65上に誤ったデータを結合するだろう。しかしながら、インバータ92、93によって引き起こされる遅延を考慮し、図9の(e)～(h)に示されているように各インバータ92若しくは93の遅延 t_d が間隔 $t_2 \sim t_1$ に存在すると仮定すれば、この誤ったヒット信号は除去される。この場合、ゲート90は2つの高状態入力を確認するこ

左タグライン 60L	右タグライン 60R
高	高
高	低
低	高
低	低

左及び右タグの両方がヒット(ライン60L及び60Rの両方が高状態)を示した場合には、バンクはいずれも選択されない(ライン60及び60Rの両方が高状態)ことに注意すべきである。これは、同じタグが左及び右バンクの両方に記憶された状態であると共に、最もエラーを表示しやすい異常な状況である。故にミス状態は信号合図されるべきである。なぜなら、このタグに関して記憶されたデータに対する2つのキーワードは矛盾するため、いずれも使用されるべきではなく、この結果、ライン62L、62Rの両方がミスを信号合図し、ゲート63L、63Rがディスエイブルされ、キャッシュ前記/Oライン64L若しくは64Rとバス65間でデータが転送されないからである。

【0030】キャッシュフィル操作(キャッシュミスの後の)では、左若しくは右バンクはゲートトランジスタ96L、96Rのゲートに適用された配置制御ライン95によって配置され、それらのゲートはリフィル制御ライン98によってゲート制御がなされたトランジスタ97を通じてライン60L若しくは60Rのいずれかに選択的に接地する。この方法で左若しくは右バンクのいずれかがディスエイブルされ、ライン60L若しくは60Rをミス状態に保持し、そのバンクに対するライン62L若しくは62Rがあるバンクに対するゲート63L若しくは63Rをオン状態とすることを防止する。他のバンクに対するバンク選択は強制ヒット(forced hit)

(比較イネブルがオフ状態とされる)であるが、ヒットに対してライン62L若しくは62R上にバンク選択信号を作り出し、ゲート63L若しくは63Rの1つを通じたデータ書き込みを可能にする。

【0031】特別な実施例を参照しつつ本発明を説明し

とはない。なぜなら、ライン60L上の電圧降下が時間間隔 t_d だけ遅延されるからである。インバータによって設けられたこれらの遅延によって、ライン60L、60R上の信号に応答可能となるようバンクセクタに対するウィンドウを定めるクロックエッジを必要とするものとは異なり、自己タイミングのマージン(一方のバンクを他方のバンクに対してタイミングを図る)が可能とされる。ライン60L若しくは60R上の電圧は、トランジスタ81若しくは82の1個若しくは20個がオン状態とされたかに拘わらずほぼ同じ速度で降下するため、電流制限トランジスタ86、87はタイミング取りも目的としている。

【0029】様々な入力に対する出力62R及び62Lの状態を以下の表に述べる。

左バンク選択ライン 62L	右バンク選択ライン 62R
高	高
低	高
高	低
高	高

てきたが、この記述は限定を意味するものではない。この記述を参照すれば、本発明の他の実施例と同様に、開示した実施例の様々な変形が当業者に明かとなろう。故に、特許請求の範囲は、本発明の意図する範囲に存在するそのような変形若しくは実施例の全てをカバーするものである。

【図面の簡単な説明】

【図1】本発明の一実施例によるキャッシュ比較及びバンク選択回路を用いるコンピュータシステムを、ブロック形態で示す図。

【図2】図1のCPUのクロック発生器によって作り出され、このCPUの内部で用いられる4つの位相クロックのタイミング図を、図1のシステムの外部バスサイクル、及びこの外部バスサイクルの形成に使用されるタイミング図と共に示した図。

【図3】図1のコンピュータシステムにおける仮想及び物理アドレス、及びキャッシュアドレッシングのフォーマットを示す図。

【図4】本発明の一実施例による図1のシステムのCPUの第1キャッシュをブロック形態で示す図。

【図5】図4のキャッシュ配列を絵画的に示した図。

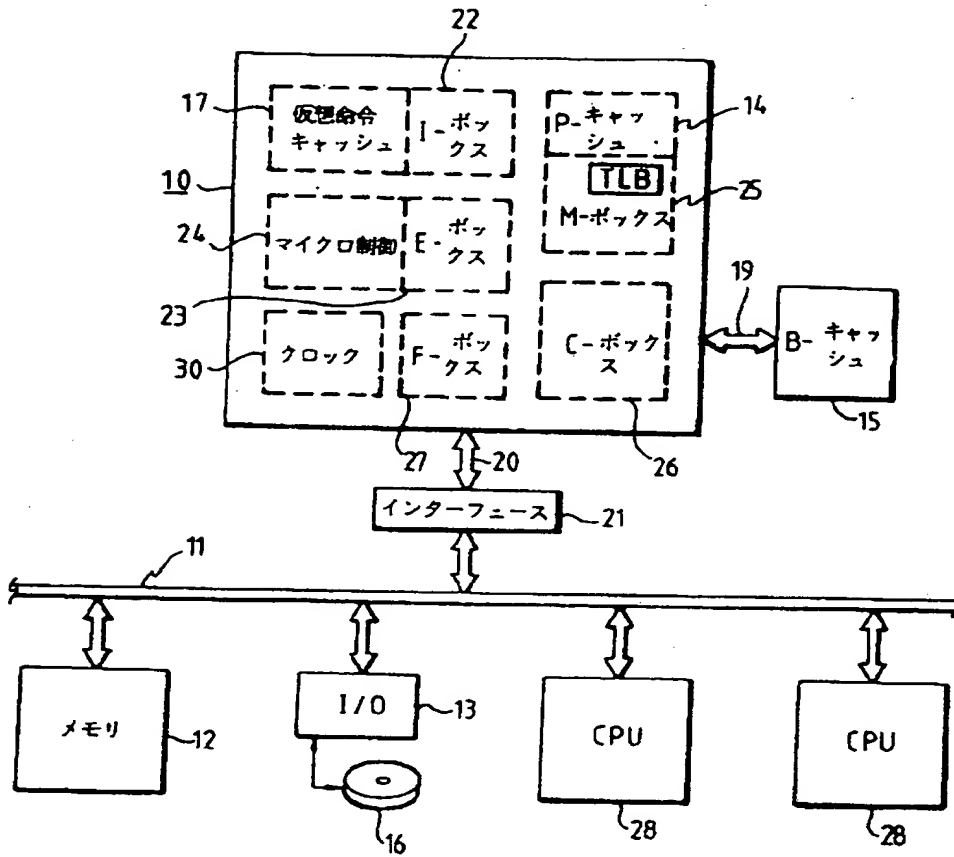
【図6】図4及び図5のキャッシュで使用されるキャッシュラインを示す図。

【図7】図4及び図5のキャッシュ動作を事象と時間との関係として示した図。

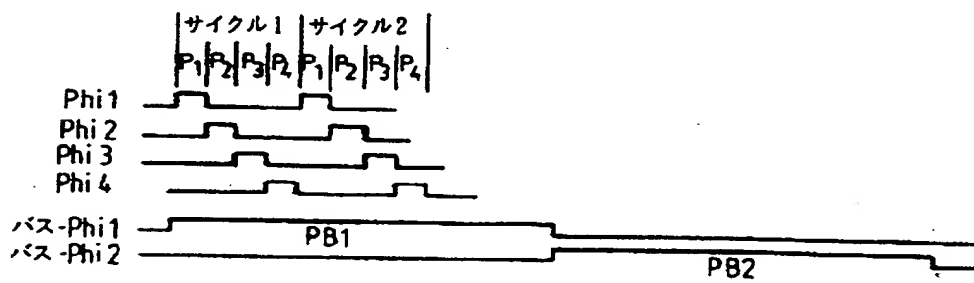
【図8】図4のキャッシュコントローラのタグ整合及びバンク選択回路を示す図。

【図9】本発明の一実施例における図8のバンク選択回路で発生する事象に関して、電圧と時間の関係を示すタイミング図。

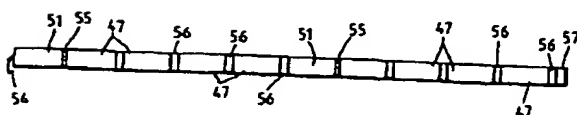
【図 1】



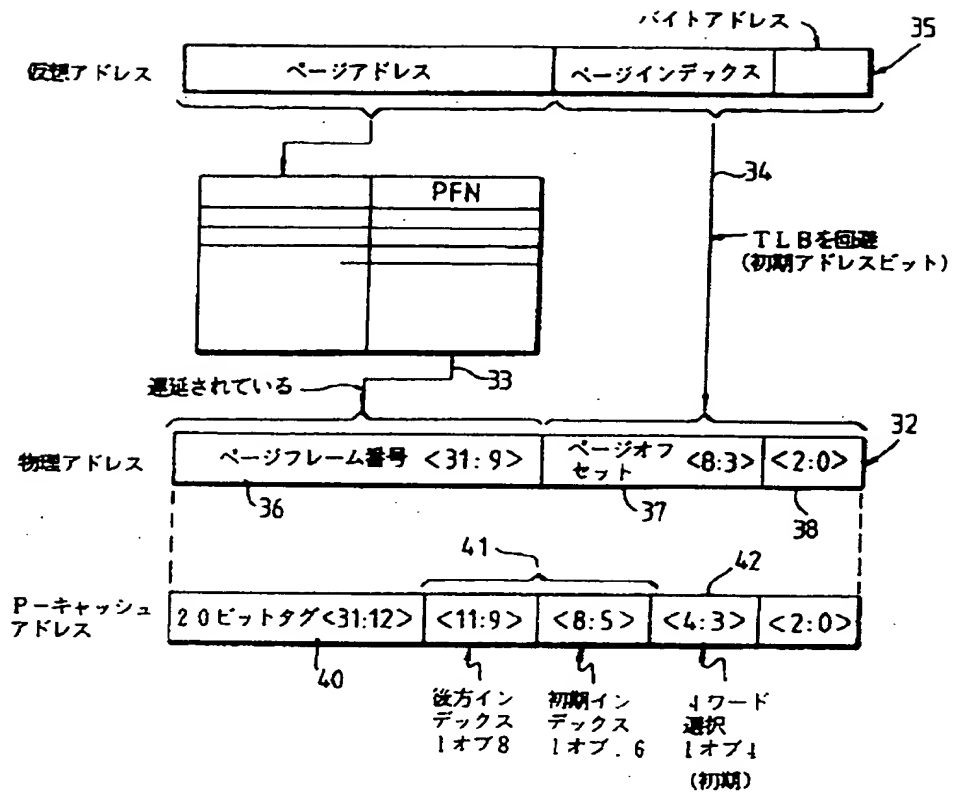
【図 2】



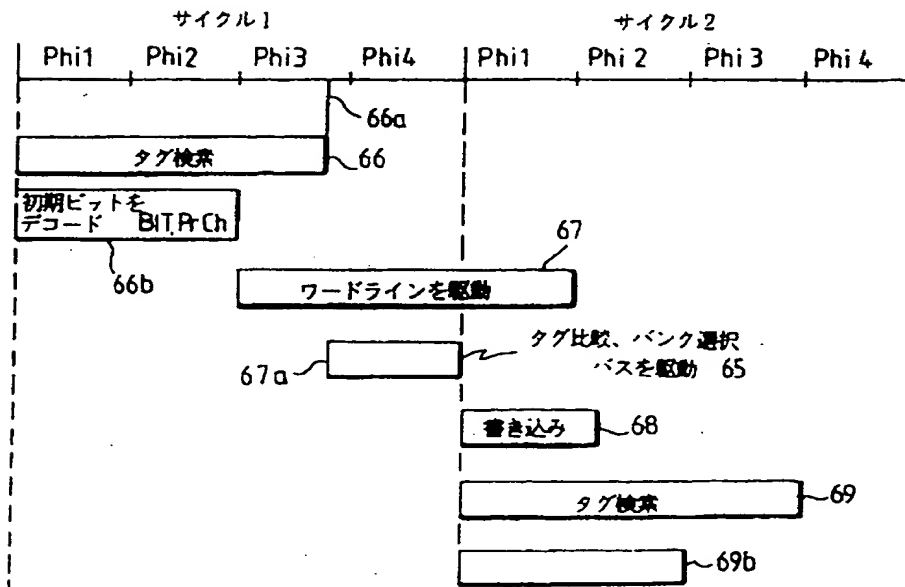
【図 6】



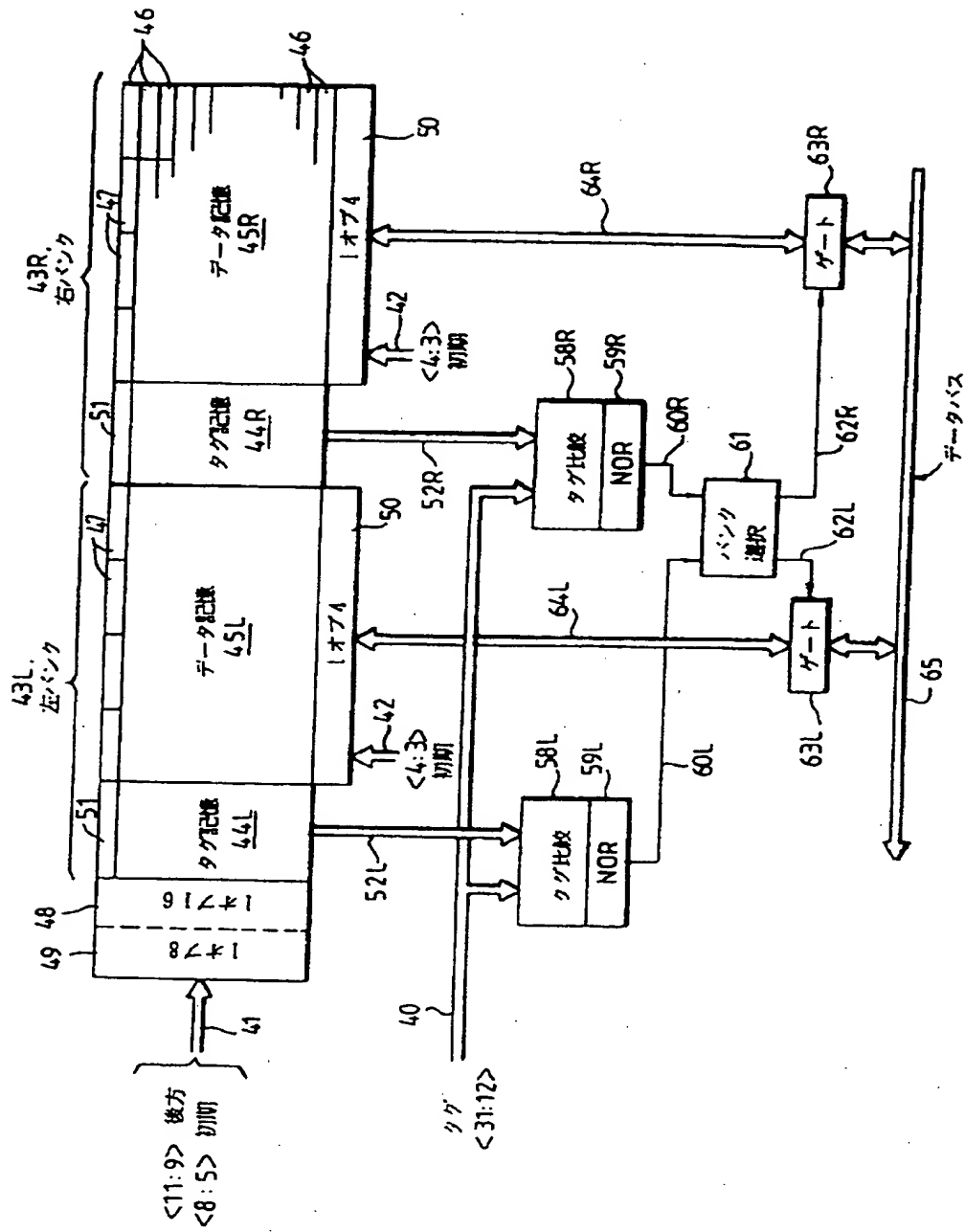
【図3】



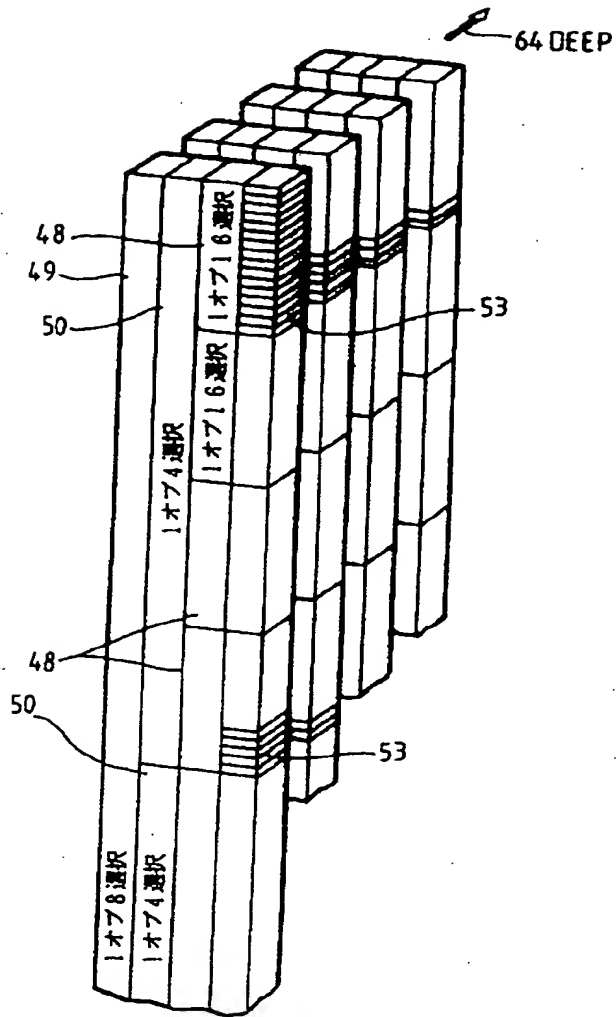
【図7】



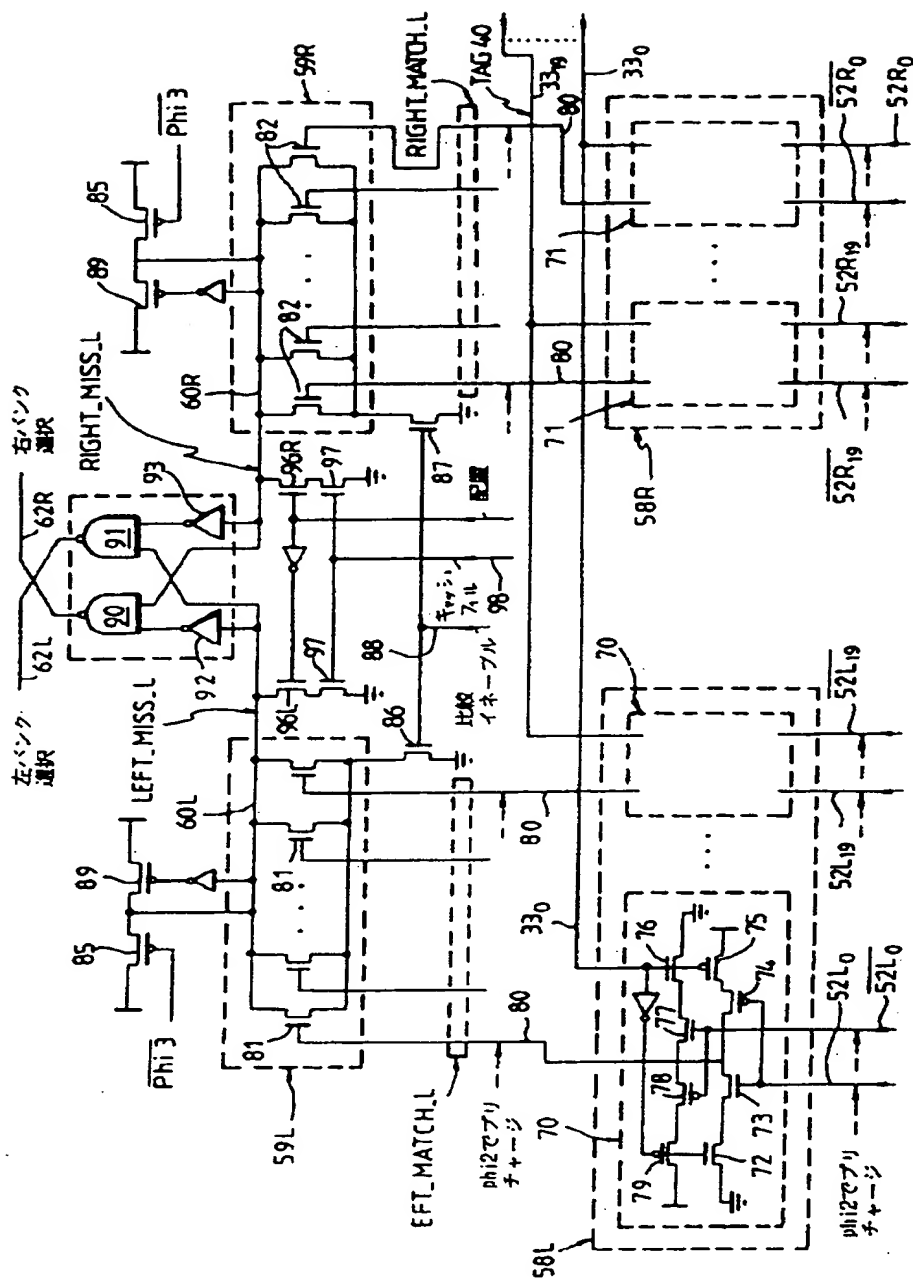
【図4】



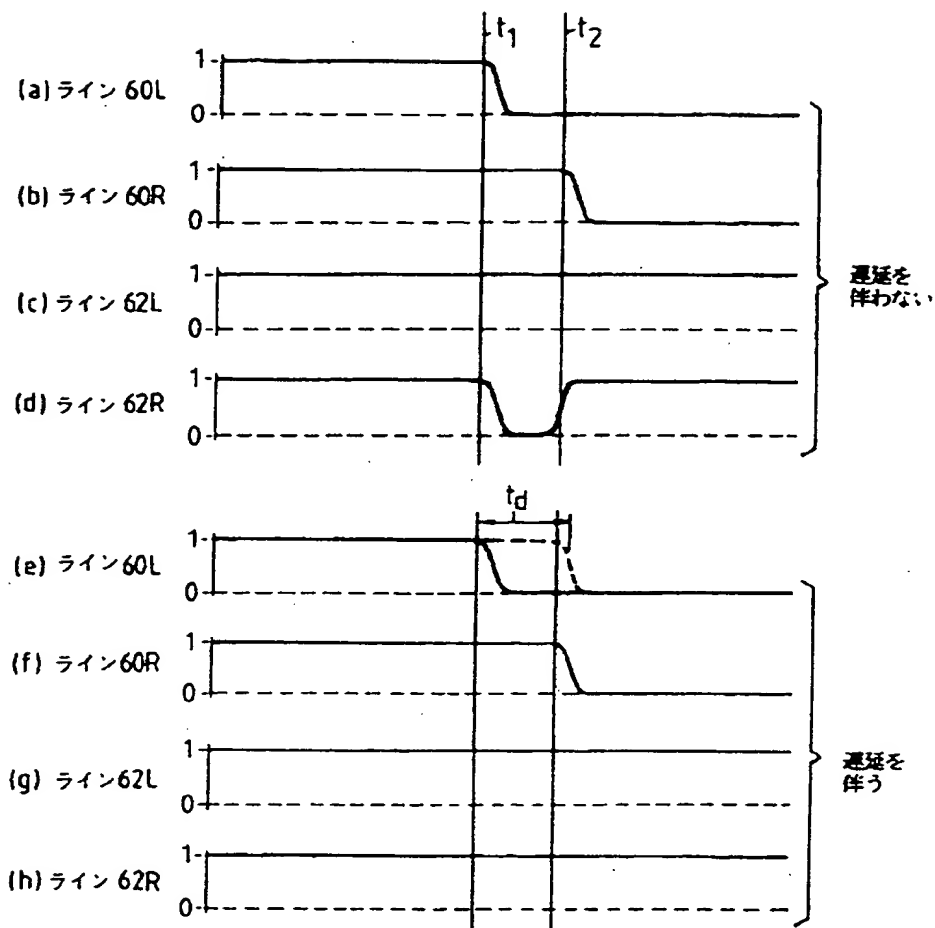
【図5】



【図8】



【図9】



フロントページの続き

(72)発明者 ウィリアム アール ウィーラー
アメリカ合衆国 マサチューセッツ州
01749ハドソン ワシントン ストリート
48

(72)発明者 マイケル リアリー
アメリカ合衆国 マサチューセッツ州
01522ジェファーソン メイソン ロード
111

(72)発明者 マイケル エイ ケース
アメリカ合衆国 マサチューセッツ州
01519グラフトン ゴードン サークル
41

(72)発明者 スティーヴン バトラー
アメリカ合衆国 マサチューセッツ州
01752マールボロ ブライアウッド レー
ン 41-7

(72)発明者 ラジス カンナ
アメリカ合衆国 マサチューセッツ州
01581ウェストボロ ウィンザー リッジ
ドライブ 2012

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.